Spatial Econometrics: A Broad View

Giuseppe Arbia
Università Cattolica del Sacro Cuore, Rome, Italy
giuseppearbia13@gmail.com
Editorial Scope

Topics

Foundations and Trends® in Econometrics publishes survey and tutorial articles in the following topics:

- Econometric models
- Simultaneous equation models
- Estimation frameworks
- Biased estimation
- Computational problems
- Microeconometrics
- Treatment modeling
- Discrete choice modeling
- Models for count data
- Duration models
- Limited dependent variables
- Panel data
- Time series analysis
- Latent variable models
- Qualitative response models
- Hypothesis testing
- Econometric theory
- Financial econometrics
- Measurement error in survey data
- Productivity measurement and analysis
- Semiparametric and nonparametric estimation
- Bootstrap methods
- Nonstationary time series
- Robust estimation

Information for Librarians

Foundations and Trends® in Econometrics, 2015, Volume 8, 4 issues. ISSN paper version 1551-3076. ISSN online version 1551-3084. Also available as a combined paper and online subscription.

Full text available at: http://dx.doi.org/10.1561/0800000030
Spatial Econometrics: A Broad View

Giuseppe Arbia
Università Cattolica del Sacro Cuore, Rome, Italy
giuseppearbia13@gmail.com
# Contents

1 Introduction 3

2 The Spatial Econometrics of Regional Data 8  
2.1 The spatial weight matrix and Moran’s I test 8  
2.2 Homoscedastic isotropic linear regression models 11  
2.3 Heteroscedastic isotropic linear regression models 35  
2.4 Non-linear isotropic models for binary response variables 38  
2.5 Homoscedastic anisotropic linear models 47

3 The Spatial Econometrics of Individual Data 51  
3.1 Spatial microeconometrics 51  
3.2 Special features of spatial microeconometrics 55  
3.3 Microdata quality issues 57  
3.4 Spatial econometrics linear models for spatial microdata 59  
3.5 Modelling the individual location of economic agents 68

4 The Spatial Econometrics of Interaction Data 82  
4.1 The spatial interaction model 82  
4.2 Spatial interaction and network dependence 83  
4.3 Modelling flows in a spatial regression 85  
4.4 Non-sample information and the zero flow problem 88  
4.5 Future challenges of spatial interaction modelling 89
Abstract

Spatial econometrics can be defined in a narrow and in a broader sense. In a narrow sense it refers to methods and techniques for the analysis of regression models using data observed within discrete portions of space such as countries or regions. In a broader sense it is inclusive of the models and theoretical instruments of spatial statistics and spatial data analysis to analyze various economic effects such as externalities, interactions, spatial concentration and many others. Indeed, the reference methodology for spatial econometrics lies on the advances in spatial statistics where it is customary to distinguish between different typologies of data that can be encountered in empirical cases and that require different modelling strategies. A first distinction is between continuous spatial data and data observed on a discrete space. Continuous spatial data are very common in many scientific disciplines (such as physics and environmental sciences), but are still not currently considered in the spatial econometrics literature. Discrete spatial data can take the form of points, lines and polygons. Point data refer to the position of the single economic agent observed at an individual level. Lines in space take the form of interactions between two spatial locations such as flows of goods, individuals and information. Finally data observed within polygons can take the form of predefined irregular portions of space, usually administrative partitions such as countries, regions or counties within one country.

In this monograph we will adopt a broader view of spatial econometrics and we will introduce some of the basic concepts and the fundamental distinctions needed to properly analyze economic datasets observed as points, regions or lines over space. It cannot be overlooked the fact that the mainstream spatial econometric literature was recently the subject for harsh and radical criticisms by a number of papers. The purpose of this monograph is to show that much of these criticisms are in fact well grounded, but that they lose relevance if we abandon the narrow paradigm of a discipline centered on the regression analysis of regional data, and we embrace the wider acceptation adopted here. In Section 2 we will introduce methods for the spatial econometric analysis of regional data that, so far, have been the workhorse of
most theoretical and empirical work in the literature. We will consider modelling strategies falling within the general structure of the SARAR paradigm and its particularizations by presenting the various estimation and hypothesis testing procedures based on Maximum Likelihood (ML), Generalized Method of Moments (GMM) and Two-Stage Least Squares (2SLS), that were proposed in the literature to remove the inefficiencies and inconsistencies arising from the presence of various forms of spatial dependence. Section 3 is devoted to the new emerging field of spatial econometric analysis of individual granular spatial data sometimes referred to as spatial microeconometrics. We present modelling strategies that use information about the actual position of each economic agent to explain both individuals’ location decisions and the economic actions observed in the chosen locations. We will discuss the peculiarities of general spatial autoregressive model in this setting and the use of models where distances are used as predictors in a regression framework. We will also present some point pattern methods to model individuals’ locational choices, as well as phenomena of co-localization and joint-localization. Finally in Section 4 the general SARAR paradigm is applied to the case of spatial interaction models estimated using data in the form of origin–destination variables and specified following models based on the analogy with the Newtonian law of universal gravitation. The discussion in this monograph is intentionally limited to the analysis of spatial data observed in a single moment of time leaving out of presentation the case of dynamic spatial data such as those observed in spatial panel data.

DOI: 10.1561/0800000030.
Spatial econometrics is still a relatively young discipline in the wider scenario of the scientific thought. Indeed, the term “spatial econometrics” was introduced for the first time 40 years ago by the Belgian economist Jean Paelinck (universally recognised as the father of the discipline) in the general address he delivered at the annual meeting of the Dutch Statistical Association in May 1974 [see Paelinck and Klaassen, 1979]. However, even if the discipline can be considered still in its adolescence compared with the more adult field of econometrics (almost 50 years older), its childhood and adolescence were anything but serene, being continuously agitated by a series of problems and challenges related to the evolution of widespread computer technologies in the 1980s, to the development of the New Economic Geography theories in the 1990s and, finally, to the explosion of the Big Spatial Data revolution in the last decade. The interest in the discipline had a particularly dramatic improvement in the last two decades which recorded an incredible explosion in the number of applied disciplines interested in the subject and, consistently, of the number of papers that appeared in scientific journals. In a comprehensive review which appeared a few years ago, Arbia [2011] surveyed 237 papers devoted
to the subject that were published only from 2007 to 2011 with an accelerated increasing trend. In trying to identify at least the major application fields, we can mention subjects such as regional economics, criminology, public finance, industrial organization, political sciences, psychology, agricultural economics, health economics, demography, epidemiology, managerial economics, urban planning, education, land use, social sciences, economic development, innovation diffusion, environmental studies, history, labour, resources and energy economics, transportation, food security, real estate, marketing and many others. But the list of applied disciplines that can benefit from the advances in spatial econometrics is, in fact, a lot longer and likely to further increase in the future.

Spatial econometrics can be defined in a narrow and in a broader sense.

In a narrow sense it refers to methods and techniques for the analysis of regression models using data observed within discrete portions of space. In a broader sense it is inclusive of the models and theoretical instruments of spatial statistics and spatial data analysis to analyze various economic effects such as externalities, interactions, spatial concentration and many others. Indeed, in the already quoted foundering book by Paelinck and Klaassen [1979] the authors indicate four paradigmatic models that constitute the backbone of the discipline. The first model is a spatial income-generating model revised to account for the presence of spatial spillovers. The model is basically an asymmetric simultaneous spatial autoregressive model [Whittle, 1954], that, for three regions, takes the form:

\[
\begin{align*}
y_1 &= \alpha^\text{au} y_2 + \beta^\text{au} y_3 \\
y_2 &= \alpha^\text{ua} y_1 + \beta^\text{uu} y_3 \\
y_3 &= \alpha^\text{ua} y_1 + \beta^\text{uu} y_2
\end{align*}
\]

with \(y_i\) indicating the income of region \(i\), \(\alpha^\text{au} \neq \alpha^\text{ua}\), \(\beta^\text{au} \neq \beta^\text{ua}\) and the superscripts \(a\) and \(u\) indicating, respectively, an agricultural and an urban area. The second paradigmatic model described by Paelinck and Klaassen [1979] is the attraction model, an example of sectorial Weberian–Loeschan location analysis [Weber, 1909; Loesch, 1940].
which takes the form of a system of equations to explain the flows of import, export and transportation costs between regions. The third model was termed the threshold effect, a mechanism that relates the birth–growth–death of new establishments to regional location profiles that, in turn, are indicative of the expected relative profitability of production in each region. Finally the fourth model is a space-integrating shopping model that aims at explaining the flows of customers towards a set of spatially located shopping centers. In the same book Paelinck and Klaassen [1979] identify five fundamental characteristics of spatial econometrics, namely: (i) the role of interdependence in spatial models; (ii) the asymmetry of spatial relations; (iii) the importance of explanatory factors located in other spaces; (iv) the differentiation between ex-post and ex-ante interaction; and (v) the explicit modelling of space. Many of these characteristics have been rather overlooked in the narrower interpretation of spatial econometrics.

As a matter of fact, the fathers of the discipline were very well clear in mind that spatial econometrics should be concerned with a range of economic and statistical modelling framework much wider than the simple regression analysis applied to regional data. Indeed, in general terms, the reference methodology for spatial econometrics lies on the advances in spatial statistics, a branch of statistics that, founded on the mathematical theory of random fields [Yaglom, 1957, 1962, Guyon, 1995, Gaetan and Guyon, 2010], suggests possible analytical solutions to the problems raised by the analysis of spatial data [Whittle, 1954, Besag, 1974, Cliff and Ord, 1972, 1981, Ripley, 1981, Cressie, 1993, Shabenberger and Gotway, 2005]. In spatial statistics it is customary [e.g. Cressie, 1993] to clearly distinguish between different typologies of data that can be encountered in empirical cases and that require different methodological approaches. A first distinction is between continuous spatial data and data observed on a discrete space. Examples of continuous spatial data are very common in many scientific fields such as physics, meteorology and environmental studies: the level of SO2 concentration, the temperature or altitude are good examples of data that can be observed, at last in principle, on a two- (or even three-) dimensional continuous surface. Such typology of data is dominant in
Introduction

physical geography, and can sometimes be relevant also in the economic analysis, for instance when studying hedonic prices in the house market. Discrete spatial data can manifest themselves in three different forms, namely points, lines and polygons. Point data are very common in the economic analysis and refer to the single economic agent (household, firm etc.) observed at an individual microlevel. Lines in space take the form of interaction flows between two spatial locations such as shopping trips, journey-to-work, communication or transportation flows and so on. Finally data observed in polygons can take the form of either regular areas (such as squares in a regular lattice grid constituting a satellite or a computer image) or predefined irregular portions of space, usually administrative partitions such as countries, regions or counties within one country. It is clear that all these typologies of spatial data are potentially of interest in spatial econometrics.

In this paper, faithful to the spirit of Paelinck and Klaassen [1979] and following the general spatial statistical approach (and also the interpretation given by the by-law of the Spatial Econometrics Association; see SEA, 2006), we will adopt this broader view of spatial econometrics [see also Griffith and Paelinck, 2011]. Consequently we will introduce some of the basic concepts and fundamental distinctions needed to properly analyze economic datasets observed on points, regions and lines. We leave intentionally out of the discussion the models designed for continuous spatial data that are still not considered in the spatial econometric literature (for one remarkable exception see Banerjee, 2015) and whose analysis would require a thorough presentation that goes beyond the limits of the present paper. The interested reader is referred to Shabenberger and Gotway [2005] and Arbia [2017] for details.

It cannot be overlooked the fact that the mainstream spatial econometric literature was recently the subject to harsh and radical criticisms by a number of papers [Partridge et al., 2012, Corrado and Fingleton, 2012, Pinkse and Slade, 2010, Gibbons and Overman, 2012]. In particular, Pinkse and Slade [2010] identify a number of methodological problems in spatial econometrics emerging mainly from the endogeneity induced by the presence of missing data and uncertainty
of location. Gibbons and Overman [2012] go further in their criticism with a paper with a provocative title which asserts that spatial econometrics is “mostly pointless” because identification problems bedevil most applied studies. The purpose of this paper is to show that much of these criticisms are in fact well grounded, but that they lose relevance if we abandon the (alas, still prevailing!) narrow paradigm of a discipline centered on the regression analysis of regional data, and we embrace the wider definition originally proposed by Paelinck and Klaassen [1979].

It is, finally, necessary to clarify that in this monograph we will limit ourselves to synchronic spatial data observed in a single moment of time. The analysis of dynamic spatial data, as those observed in spatial panels, is left out of the range of interest of this monograph. We can, however, rely on the comprehensive paper by Lee and Yu [2011] published in a previous issue of Foundations and Trends in Econometrics which treats the topic in details and to which we are happy to refer the interested readers.
References


References


G. Arbia, M. Bee, G. Espa, and F. Santi. Fitting spatial regression to large datasets using unilateral approximations. working paper, University of Trento, 2015a.

References


References


References


References


References


References


